AD-A115 312 STANFORD UNIV CA F/G 12/1

A SAMPLING PROCEDURE AND PUBLIC POLICY.(U)

MAR 82 G J LIEBERMAN, I OLKIN, F RIDDLE N00014-75-C-0561

UNCLASSIFIED TR-203 NL

1 OF 1

AD A

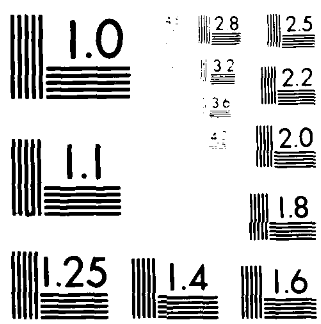
115312

END

DATE FILMED

07-82

DTIC



MICROCOPY RESOLUTION TEST CHART

AD A115312

A SAMPLING PROCEDURE AND PUBLIC POLICY

by

G. J. LIEBERMAN
I. OLKIN
and
F. RIDDLE

TECHNICAL REPORT NO. 203

March 1982

SUPPORTED UNDER CONTRACT N00014-75-C-0561 (NR-047-200)
WITH THE OFFICE OF NAVAL RESEARCH

Gerald J. Lieberman, Project Director

DEPARTMENT OF OPERATIONS RESEARCH
AND
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

A SAMPLING PROCEDURE AND PUBLIC POLICY

G.J. Lieberman, I. Olkin, F. Riddle
Stanford University

Acknowledgement

The authors wish to acknowledge a number of conversations with colleagues concerning the concept of statistical sampling in this context. In particular, we are exceedingly grateful to Peter Bacchetti who helped immeasurably with the details and computations of the sampling plan.

Gerald J. Lieberman is currently Vice Provost and Dean of Graduate Studies and Research and is a member of the Departments of Statistics and Operations Research at Stanford University; Ingram Olkin is Professor of Statistics and Education at Stanford University; Franklin Riddle is Associate Controller at Stanford University.

A Sampling Procedure and Public Policy

G. J. Lieberman, I. Olkin, F. Riddle

Stanford University

Following the March, 1979 publication of the revised Office of Management Budget Circular A-21 "Cost Principle for Higher Education", which promulgates the cost principles to be used in the determination of cost charged to government sponsored agreements, there was an undercurrent of faculty dissatisfaction with the revised rules. Of major issue was, and still is, the requirements relating to reporting of faculty effort charged directly or indirectly to government agreements.

The two acceptable methods for effort reporting contained in the revised A-21 are contained in section J.6. of that document and are called Monitored Workload and Personal Activity Reporting. The Monitored Workload method can be used for Faculty and Professionals only, whereas the Personal Activity Reporting method can be used for all personnel.

The major objection raised by the faculty centered on the requirement (under either system) to account for 100% of their effort applicable to each sponsored agreement, each indirect cost category, and each major function of the University. This level of accountability was considered burdensome, arbitrary, and to some extent, useless work because of the interrelated nature of the research, instruction and other activities of the faculty. The results of such effort reports, though fulfilling the documentation requirements established by the regulation, were known to be arbitrary and therefore inaccurate, and with the added detail terribly expensive for institutions to implement.

That the assessment of direct and indirect costs is, at times, highly subjective and approximate was also recognized by the government. Section J.6.b of OMBA-21 states:

> ". . . it is recognized that, because of the nature of the work involved in academic institutions, the various and often interrelated activities of professional and professional employees frequently cannot be measured with a high degree of precision, that reliance must be placed on reasonably accurate approximations and that acceptance of a degree of tolerance in measurement is appropriate."

Because of dissatisfaction with the new effort reporting guidelines, the major research institutions formed a task force under the auspices of the Association of American Universities (AAU) to develop alternative methods of gathering faculty effort without the burden necessitated by the new regulations. As a consequence of their deliberations, it was suggested that a method of statistical sampling be used to determine the indirect effort performed by the faculty. The results of the sample could then be used in determining indirect cost rates.

It was also noted that too often the approximations made in effort reporting resulted in very poor data. To improve the data, the faculty would have to be apprised of what specific costs are to be labeled a direct or indirect cost. By cutting the scope of reporting, a sampling procedure would permit increased dialogue between the administration and faculty thereby enabling the faculty to have a better understanding of the process.

The idea of statistical sampling was explored between Stanford University and the Office of Management and Budget. The OMB felt the idea had possibilities and suggested that Stanford submit a formal proposal. The authors carried out some preliminary work to describe how such a system would work, and a plan was presented to OMB and a panel of Federal statisticians. After their review, agreement was reached to proceed with a more detailed study to include the process, the sample sizes to reach a 90% confidence level with a probable 5% error and the statistical formulas to produce the desired data.

The plan was completed in January, 1981 and submitted and accepted by OMB which had chosen ten universities who had expressed interest in using the statistical method as test of the proposal. Unfortunately, of those institutions who submitted their proposals based on the statistical model, the implementing agency wanted the test made in parallel with a standard J.6. method so that the results could be verified. In essence, this stringent requirement virtually terminated further consideration by the institutions.

However, on January 7, 1982, after several months of added effort by the AAU and other higher education associations, a proposal for revision of section J.6. of A-21 was placed in the Federal Register. Included in the proposed revision is paragraph J.6.b(2)(c), which reads:

(c) The payroll distribution system will allow confirmation of activity allocable to each sponsored agreement and each category of activity needed to identify indirect costs and the functions to which they are applicable. The activities chargeable to indirect cost categories or the major functions

> of the institution for employees whose salaries must be apportioned (see J.6.b(1)(b) above) if not initially identified as separate categories, may be subsequently distributed by any reasonable method mutually agreed to, including, but not limited to, suitably conducted surveys, statistical sampling (emphasis added) procedures, or the application of negotiated fixed rates.

It is assumed that the public comment will be overwhelmingly favorable to the proposed revision, which makes the publication of the method developed at Stanford by the authors timely. Use of the method would allow confirmation of salaries directly charged to sponsored agreements by confirming notations on the monthly expenditure reports and less costly (and less disruptive for the faculty), gathering of data necessary for indirect costing purposes. What follows is the proposal that was submitted to OMB as a statistical sampling model.

2. The Sampling Plan

The proposed sampling plan is designed for a particular university for a particular purpose. However, the method can be extended to other organizations and for other purposes.

The sample is designed to determine the amount of effort expended by the faculty in indirect activities such as departmental research, university and student service administration. For example, if the indirect activity departmental administration is of interest, then we need to estimate the total cost, T, expended by the faculty engaged in that activity:

$$T = f_1 s_1 + \cdots f_N s_N \quad ,$$

where f_j is the fraction of time that the j-th faculty members devotes to departmental administration, s_j is the academic year salary of the j-th faculty member and N is the number of relevant faculty. There will be similar totals for university administration, research administration, etc.

Because faculty differ with respect to their activities, we use a stratified sample thereby reducing the sample variance of our estimates. Two methods of stratification appear natural in a university context, namely, rank and disciplines.

For stratification by rank four strata may be appropriate: professors, associate professors, assistant professors, department chairs and faculty with major administrative duties. Stratification by discipline might include the Schools of Medicine, Engineering, and Education, the Physical Sciences, Social Sciences, the Humanities, etc. In making these divisions it is important to emphasize that within each stratum the faculty should be relatively homogeneous.

Definition of the Universe

The universe is assumed to be all faculty who conduct sponsored activities and have other duties funded from non-government sources, including the departmental operating budget. Also included would be faculty totally funded from the operating budget or a combination of operating budget and other non-government sources. A faculty person funded 100% on sponsored agreements or patient care would be excluded from the universe. Since the results of the sample needs to be generalized to the universe it represents, the sample must reflect the elements necessary to fulfill the objective of the universe.

The Strata

At Stanford University strata were chosen using a combination of both the rank and discipline criteria. These were:

1) Medical School department chairs.
2) Other department chairs and laboratory directors.
3) Full professors in the sciences.
4) Full professors in other fields.
5) Associate professors.
6) Assistant professors.

3. Details of the Sampling Plan

3.1 Preliminaries. For each stratum, we have available the total number, N_i, of faculty members, and the average, $\bar{Y}_i$, of the product of academic year salary and proportion of time devoted to departmental administration. Consequently, we can also determine the sample variance, S_i^2, of the Y-values. That is

$$S_i^2 = \Sigma(Y-\bar{Y}_i)^2/(N_i-1) \quad .$$

Actually, much of the above data is unknown beforehand. However, some preliminary estimates are essential if we are to sample with a specified degree of accuracy. One procedure is to obtain estimates from past data, or from a sample of past data. Indeed, an examination of past data may also lead to improved stratification. At Stanford an examination of 1979-1980 data was used to estimate the needed parameters and also to establish the use of the strata.

3.2 Choice of Sample Sizes. The design depends on choosing an overall sample size, n, and an allocation into the various strata. The total sample size depends on the confidence level, c, and a given relative error, e. The higher the required confidence, the larger the required sample size; the smaller the relative error, or equivalently, the greater the precision, the larger the required sample size. The sample size also depends on the variances in each stratum, but not in a simple way.

The determination of n is given by

$$n \cong \frac{(\Sigma N_i S_i)^2}{W^2 + \Sigma N_i S_i^2} = \frac{T^2}{W^2 + V} \quad , \tag{1}$$

where $T = \Sigma N_i S_i$, $V = \Sigma N_i S_i^2$ and W depends on the confidence level, c, and the relative error e:

$$W = e(\text{Total})/c^* \quad , \tag{2}$$

where the total is the estimated true total cost of administration, $c^* = \Phi^{-1}((1+c)/2)$, and Φ is the standard normal cumulative function.

Once the sample size is chosen, the allocation n_i to stratum i is made in the proportion

$$n_i = \left(\frac{N_i S_i}{\Sigma N_j S_j}\right) n, \qquad n_i \leq N_i \quad . \tag{3}$$

3.3 An Example

The following data with 5 strata are fictitious.

Stratum	Number N_i in stratum	Variance S_i^2 in each stratum (in millions)	Stratum standard deviation S_i (in thousands)	$N_i S_i$ (thousands)	$\frac{N_i S_i}{\Sigma N_j S_j}$	$N_i S_i^2$ (in millions)
A	B	C	D	E	F	G
1	20	100.00	10.0	200	.110	2000
2	50	4.00	2.0	100	.055	200
3	200	9.00	3.0	600	.329	1800
4	250	6.25	2.5	625	.342	1562.5
5	300	1.00	1.0	300	.164	3.0
				T = 1,825	1.000	V = 5,862.5

If we wish to estimate the total amount spent on departmental administration to within 5% of the true total with 90% confidence, and if we estimate the true total to be more than \$7 million, then from (2)

$$W = \frac{(.05)(7,000,000)}{1.645} = 212,766 \ ,$$

where $c^* = 1.645$ is obtained from the tables of normal distribution. From (1)

$$n \geq \frac{(1,825,000)^2}{(212)^2 + (5,862,500,000)} \approx 65.1 \quad ,$$

which we round upward to $n = 66$.

The allocation to strata is now obtained from the proportions in column F. Thus, stratum 1 will have a sample size $66(.11) = 7.3$. The results for the five strata are

$$7.3, \quad 3.6, \quad 21.7, \quad 22.6, \quad 10.8 \; .$$

Since these values are not integers, it will be safest to round up to

$$8, \quad 4, \quad 22, \quad 23, \quad 11 \; ,$$

which yields a total sample size of 68.

Because Stanford is on a quarter system, the sample sizes within strata are divided into thirds, representing the three quarters. To avoid unequal numbers per quarter, and to simplify the procedure, we round upwards so that the samples are divisible by 3, namely,

$$9, \quad 6, \quad 24, \quad 24, \quad 12 \; .$$

Thus, for example, we would sample 4 per quarter in the fifth stratum. The total sample size would now be 75.

<u>Remark</u>. An alternative procedure, if a sample is not divisible by 3, is to randomize the remainder. For example, in the fifth stratum, take 3 faculty per quarter and assign the remaining two at random to two quarters.

4. Derivations and Technical Results

We now provide some of the derivations and rationale for the formulas. Formulas that are standard are not derived.

Let y_{ijk} be the y value of the j-th faculty member in stratum i for quarter (semester) k. [Since Stanford is on a quarter system, we use 3 quarters in our discussion. However the procedures are easily modified for use at a university with the semester system.]

A sample of size n is chosen and then partitioned in accordance with the allocation

$$n_i = \frac{N_i S_i}{T} n , \qquad n_i \le N_i ,$$

in each stratum, where we assume that n_i is rounded upward to be divisible by 3 (or 2). We then measure the y-value of 1/3 (or 1/2) of the sample in each stratum each quarter (or semester).

A notation is needed to denote measurements taken in each of the quarters, and we use $\dot{y}_i$, $\ddot{y}_i$ and $\dddot{y}_i$ to denote the sums of y-values in quarters 1, 2, 3, respectively. That is

$$\dot{y}_i = \sum_{j=1}^{\bar{n}_i} y_{ij1} ,$$

$$\ddot{y}_i = \sum_{j=\bar{n}_i+1}^{2\bar{n}_i} y_{ij2}$$

$$\dddot{y}_i = \sum_{j=2\bar{n}_i+1}^{n_i} y_{ij3} ,$$

where $\bar{n}_i = n_i/3$.

The sampling procedure has two-stages: In the first stage we sample the faculty within each stratum, and in the second stage, sample by quarters (semesters). The quantity to be estimated is

$$Y \equiv \sum_{i=1}^{g} \sum_{j=1}^{N_i} \sum_{k=1}^{3} Y_{ijk}/3 \quad , \tag{3}$$

where g is the number of strata. The estimate used is

$$\hat{Y} = \sum_i \frac{N_i}{n_i} (\dot{y}_i + \ddot{y}_i + \dddot{y}_i) \quad , \tag{4}$$

which has variance

$$V(\hat{Y}) = V(E(\hat{Y}|\text{sample})) + E(V(\hat{Y}|\text{sample})) \quad . \tag{5}$$

Presumably, $V(E(\hat{Y}|\text{sample}))$ is large compared to $E(V(\hat{Y}|\text{sample}))$.

Consider the population with values

$$y^*_{ij} = \frac{1}{3}(y_{ij1} + y_{i12} + y_{ij3}) \quad .$$

The total Y^* is just Y, and our sampling scheme reduces to a single-stage, stratified simple random sample. Also,

$$\hat{Y}^* = \sum_i \frac{N_i}{n^*_i} \sum_{j=1}^{n_i} y^*_{ij} \quad ,$$

which has variance

$$\sum_i \frac{N_i^2 (S^*_i)^2}{n^*_i} \left(1 - \frac{n^*_i}{N_i}\right) = V(\hat{Y}^*) \quad .$$

With optimal allocation, $n^*_i = n(N_i S^*_i/T^*)$, where $T^* = \Sigma N_\alpha S^*_\alpha$. Then

$$V(\hat{Y}^*) = \sum_i \frac{N_i^2 S^{*2}_i}{n \frac{N_i S^*_i}{T^*}} \left(1 - \frac{\frac{N_i S^*_i}{T^*} n}{N_i}\right) = \sum_i \frac{T^* N_i S^*_i}{n} \left(1 - n \frac{S^*_i}{T^*}\right)$$

$$= \sum_i N_i S^*_i \left(\frac{T^*}{n} - S^*_i\right) = \frac{T^{*2}}{n} - V^* \quad ,$$

where $V^* = \Sigma N_i S_i^{*2}$. Now, $S_i^* \leq S_i$, so that

$$V(\hat{Y}^*) \leq \frac{T^2}{n} - V \quad .$$

Also, because the expectation is linear,

$$V(E(Y|\text{sample})) = V(\sum_i \frac{N_i}{n_i} \sum_{j=1}^{n_i} y_{ij}^*) = V(\hat{Y}^*) \quad .$$

Consequently $V(\hat{Y}) \geq V(\hat{Y}^*)$, and

$$V(\hat{Y}) \doteq V(Y^*) \leq \frac{T^2}{n} - V \quad ,$$

which suggests the approximation

$$\frac{T^2}{n} - V \doteq V(\hat{Y}) \quad .$$

For relative error e with confidence coefficient c, we must have

$$P\{Y(1-e) \leq Z \leq Y(1+c)\} \geq c \quad ,$$

where Z is normally distributed with mean Y and variance $V(Y) \doteq (T^2/n)-V$. This requires that the condition

$$2\Phi(W) \equiv \Phi(cY/\sqrt{V(Y)}) - 1 \geq c \tag{7}$$

be satisfied. But (7) is equivalent to

$$eY/\sqrt{V(Y)} \geq \Phi^{-1}(\frac{1+c}{2}) \quad ,$$

where Φ is the standard normal cumulative distribution. Consequently,

$$V(\hat{Y}) \doteq \frac{T^2}{n} - V \leq W^2 \quad ,$$

from which we obtain

$$n \geq \frac{T^2}{W^2 + V} \quad . \tag{8}$$

This formula is slightly optimistic because it allows for fractional allocation to strata and because $V(\hat{Y})$ was approximated. Since T and V are not known exactly, further caution whould be used in picking n.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 203	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A SAMPLING PROCEDURE AND PUBLIC POLICY		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) G. J. Lieberman, I. Olkin and F. Riddle		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0561
9. PERFORMING ORGANIZATION NAME AND ADDRESS DEPT'S OF OPERATIONS RESEARCH AND STATISTICS STANFORD UNIVERSITY STANFORD, CALIFORNIA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (NR-047-200)
11. CONTROLLING OFFICE NAME AND ADDRESS OPERATIONS RESEARCH, CODE 434 OFFICE OF NAVAL RESEARCH ARLINGTON, VIRGINIA 22217		12. REPORT DATE MARCH 1982
		13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

CIRCULAR A-21	SAMPLING PROCEDURE	EFFORT REPORTING
STATISTICS	PUBLIC POLICY	

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper describes the background of the Office of Management Budget Circular A-21 "Cost Principle for Higher Education" that describes the requirement for effort reporting. A sampling procedure is proposed as an alternative to 100% reporting.

DAT
FILM